Model selection (2)

Manel Martínez-Ramón

ECE, UNM

October 2018

The constant kernel as a parameter

At this point it is worth to remember that in almost all cases, we need to add a constant term to our model. This is, the model is constructed as

$$y = \mathbf{v}^\top \boldsymbol{\varphi}(\mathbf{x}) + b$$

where $\mathbf{v}$ is a set of linear parameters, but we prefer to write that model with the form

$$y = \mathbf{w}^\top \begin{pmatrix} \boldsymbol{\varphi}(\mathbf{x}) \\ \sqrt{k} \end{pmatrix}$$

where $\mathbf{w} = \begin{pmatrix} \mathbf{v} \\ b\sqrt{k} \end{pmatrix}$ is our usual set of linear parameters.

Then, our nonlinear transformation is actually $\begin{pmatrix} \boldsymbol{\varphi}(\mathbf{x}) \\ \sqrt{k} \end{pmatrix}$ and the dot product between two data is

$$\begin{pmatrix} \boldsymbol{\varphi}(\mathbf{x}) \\ \sqrt{k} \end{pmatrix}^{\top} \begin{pmatrix} \boldsymbol{\varphi}(\mathbf{z}) \\ \sqrt{k} \end{pmatrix} = \boldsymbol{\varphi}^{\top}(\mathbf{x})\boldsymbol{\varphi}(\mathbf{z}) + k = \mathbf{k_f}(\mathbf{x}, \mathbf{z}) + k$$

The constant kernel as a parameter

- Thus, the actual kernel has three elements:

$$\mathbf{k_y}(\mathbf{x}, \mathbf{z}) = \mathbf{k_f}(\mathbf{x}, \mathbf{z}) + \sigma_n^2 \delta(|\mathbf{x} - \mathbf{z}|) + k$$

where the constant k has been included as a parameter.

- The kernel matrix contains, then, a constant matrix.

$$\mathbf{K_y} = \mathbf{K_f} + \sigma_n^2 \mathbf{I} + k \mathbf{1}_{N,N}$$

If included as a parameter, in most situations changing the value of k becomes irrelevant.

- We can compute an eigenanalisys of $\mathbf{K_f} + \sigma_n^2\mathbf{I} + k\mathbf{1}_{N,N}$ as follows. The expression of parameters $\boldsymbol{\alpha}$ are

$$\begin{aligned}\boldsymbol{\alpha} &= \left(\mathbf{K_f} + \sigma_n^2\mathbf{I} + k\mathbf{1}_{N,N}\right)^{-1}\mathbf{y} \\ &= \left(\mathbf{Q}\left(\boldsymbol{\Lambda} + \sigma_n^2\mathbf{I}\right)\mathbf{Q}^\top + kN\mathbf{u}\mathbf{u}^\top\right)^{-1}\mathbf{y}\end{aligned}$$

where straigthforwardly $\mathbf{u}$ is the only nonzero eigenvector of $\mathbf{1}_{N,N}$, with components $N^{-1/2}$ and eigenvalue N.

- Then, using twice the matrix inversion lemma, and recalling that the last value of $\boldsymbol{\varphi}(\mathbf{x}*)$ is $\sqrt{k}$, we find

$$f(\mathbf{x}^*) = \boldsymbol{\alpha}^\top \mathbf{k}(\mathbf{x}^*) - \frac{1}{N}\mathbf{1}^\top \mathbf{y}$$

  This is, adding a constant matrix simply removes the mean of the regressors, regardless of the value of k.

The constant kernel as a parameter

- Nevertheless, in a situation where the kernel is constructed using a product of kernels as for example

$$k(\mathbf{x}, \mathbf{z}) = (k_1(\mathbf{x}, \mathbf{z}) + k_a)k_2(\mathbf{x}, \mathbf{z}) + k_b$$

the previous result is not true for k_a, and then this becomes a free parameter that we have to adjust.

In this lesson, students must be able to prove:

- That usually we need to include a constant matrix added to the kernel function to include a bias constant to the estimator.
- That this constant has the effect of removing the bias of the regressor sequence $\mathbf{y}$.
- That in some cases, we need to validate a constant term added to a kernel.